LBNL-62531 ABS
UCRL-ABS-229323

DOE JOINT GENOME INSTITUTE PRODUCTION GENOMICS FACILITY (PGF) FINISHING PIPELINE

H. SUN, A. CLUM, E. GOLTSMAN, S. LOWRY, B. FOSTER, S. TRONG, P. KALE, A.L. LAPIDUS, P. RICHARDSON

DOE JOINT GENOME INSTITUTE, WALNUT CREEK, CA

The DOE Joint Genome Institute is an integrated high-throughput sequencing facility which provides high-quality genomic sequences to the scientific community. Currently there are approximately 300 microbial genomes in the JGI production genomics facility (PGF) pipeline in Walnut Creek and to date, 160 have been completely finished. The PGF microbial finishing group is involved in improving the microbial genomic sequences that come directly out of production. The finishing team is responsible for solving misassemblies, closing gaps, improving low-quality areas in the DNA assemblies and submitting high-quality data to the annotation group which in turn submit the sequence to the public domain. The finishing pipeline at PGF starts with resolving DNA repeats using an automated pipeline developed at JGI/PGF. Low quality and gap regions in the assemblies are resolved by directed reactions mainly using primers designed by applying Autofinishing functionality of Consed and other software developed in-house. Finally, manual checks are performed to ensure a minimum consensus phred quality of 30 or above, error rate is less than 1 base per 50,000bp and 2 times coverage throughout the DNA sequence assemblies. The strategies for solving repeats, misassemblies and difficult to finish regions will be further discussed in detail in this poster.

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under contract No. DE-AC52-06NA25396.

.